

Vine copulas as a mean for the construction of high dimensional probability distribution associated to a Markov Network

Edith Kovács · Tamás Szántai

Received: date / Accepted: date

Abstract Building higher-dimensional copulas is generally recognized as a difficult problem. Regular-vines using bivariate copulas provide a flexible class of high-dimensional dependency models. In large dimensions, the drawback of the model is the exponentially increasing complexity. Recognizing some of the conditional independences is a possibility for reducing the number of levels of the pair-copula decomposition, and hence to simplify its construction (see [1]). The idea of using conditional independences was already performed under elliptical copula assumptions [17], [24] and in the case of DAGs in a recent work [2].

We provide a method which uses some of the conditional independences encoded by the Markov network underlying the variables. We give a theorem which under some graph conditions makes possible to derive pair-copula decomposition of the probability density function associated to a Markov network.

As the underlying Markov network is usually unknown, we first have to discover it from the sample data. Using our results published in [33] and [21] we will show how to derive a multidimensional copula model exploiting the information on conditional independences hidden in the sample data.

Keywords Copula decomposition · t -cherry junction tree · Markov network · Cherry-wine probability distribution · Graphical models

1 Introduction

Copulas in general are known to be useful tool for modeling multivariate probability distributions since they serve as a link between univariate marginals. Pair-copula construction introduced by H. Joe [18] is able to encode more types of

Department of Mathematics, ÁVF College of Management of Budapest
Villányi út 11-13, H-1114 Budapest, Hungary

Institute of Mathematics, Budapest University of Technology and Economics,
Műgyetem rkp. 3, H-1111 Budapest, Hungary
Tel.: +36-1-463-1298
Fax: +36-1-463-1291
E-mail: szantai@math.bme.hu

dependencies in the same time since they can be expressed as a product of different types of bivariate copulas. For solving the problem which occurs when we want to find consistent marginal copulas involved in the expression of a junction tree copula density (see [21]) we found extremely useful the concept of Regular-vine copulas. Our research in this direction was also motivated by the arising open questions in the papers published in this field, as follows.

The paper [1] calls the attention on the fact that "conditional independence may reduce the number of the pair-copula decompositions and hence simplify the construction". In this paper the importance of choosing a good factorisation which takes advantage from the conditional independence relations between the random variables is pointed out. In the present paper we give a method for finding that pair copula construction which exploits the conditional independences between the variables of a given Markov network. We also give a method for constructing Regular-vines starting from a multivariate data set.

The importance of taking into account the conditional independences between the variables encoded in a Bayesian Network (directed acyclic graph) was explored in the papers [24] and [17]. Two problems of this aspect are discussed. First when the Bayesian Network (BN) is known, some of the conditional independences taken from the BN are used to simplify a given expression of the D- or C- vine copula. Secondly the problem of reconstruction of the BN from a sample data set was formulated under the assumption that the joint distribution is multivariate normal. For discovering the independences and conditional independences between the variables in [17] are used the correlations, the conditional correlations and the determinant of the correlation matrix. In the present paper we also exploit the conditional independences encoded in a Markov network which has the advantage that we do not need to know the ordering of the random variables. We will express the conditional independences in terms of information theoretical concepts which do not need any assumption on the type of copula.

In the recent work [2] Bauer et al. are dealing with a more general case with the pair-copula constructions for non-Gaussian BN. In their paper the BN is supposed to be known. The formula of probability distribution associated to the given BN is expressed by pair-copulas. A similar idea will be used in our approach, we will transform the so called cherry-tree copula introduced in [21] into a vine copula constructed from pair copula-blocks.

The truncated Regular-vine copula is defined in [23] and [5]. In [23] an algorithm is developed for searching the "best vine". This algorithm uses the partial correlations. This paper suggested us the idea to prove a theorem which ensures the construction of the best truncated Regular-vine distribution, at a given level k . In order to find such a representation we give a greedy algorithm, which generally is a good heuristic, but if some assumptions are fulfilled the algorithm results the optimal solution.

Because the work of the present paper is strongly related to Markov networks which also need some graph theoretical concepts, copulas and the special case of Regular-vine copulas the second part of the paper is a preliminary part that contains some of the concepts we will use throughout the paper. The third part of the paper discusses under which graphical conditions of the Markov network the multivariate copula can be expressed as a junction tree copula and as a cherry-tree copula. Then we give a pair-copula construction (formula) and a Regular-vine structure (graphical structure) of the cherry-tree copula. The fourth part of

the paper presents a method for finding the cherry tree copula starting from a multivariate sample data set. In the fifth part we discuss the properties of the best fitting probability density and copula density associated to truncated R-vine. We finish the paper with some conclusions.

2 Preliminaries

In this section we introduce some concepts used in graph theory and probability theory that we need throughout the paper and present how these can be linked to each other. For a good overview see [26].

2.1 Markov Network

We first present the acyclic hypergraphs and junction trees. We then present a short reminder on Markov network. We finish this part with the multivariate joint probability distribution associated to a junction tree.

Let $V = \{1, \dots, d\}$ be a set of vertices and Γ a set of subsets of V called *set of hyperedges*. A *hypergraph* consists of a set V of vertices and a set Γ of hyperedges. We denote a hyperedge by C_i , where C_i is a subset of V . If two vertices are in the same hyperedge they are connected, which means, the hyperedge of a hypergraph is a complete graph on the set of vertices contained in it.

The *acyclic hypergraph* is a special type of hypergraph which fulfills the following requirements:

- Neither of the edges of Γ is a subset of another edge.
- There exists a numbering of edges for which the *running intersection property* is fulfilled: $\forall j \geq 2 \quad \exists i < j : C_i \supset C_j \cap (C_1 \cup \dots \cup C_{j-1})$. (Other formulation is that for all hyperedges C_i and C_j with $i < j-1$, $C_i \cap C_j \subset C_s$ for all $s, i < s < j$.)

Let $S_j = C_j \cap (C_1 \cup \dots \cup C_{j-1})$, for $j > 1$ and $S_1 = \phi$. Let $R_j = C_j \setminus S_j$. We say that S_j *separates* R_j from $(C_1 \cup \dots \cup C_{j-1}) \setminus S_j$, and call S_j separator set or shortly separator.

Now we link these concepts to the terminology of junction trees.

The junction tree is a special tree structure which is equivalent to the connected acyclic hypergraphs [26]. The nodes of the tree correspond to the hyperedges of the connected acyclic hypergraph and are called clusters, the edges of the tree correspond to the separator sets and called separators. The set of all clusters is denoted by \mathcal{C} , the set of all separators is denoted by \mathcal{S} . The junction tree with the largest cluster containing k variables is called *k-width junction tree*. A vertex which is contained in only one cluster is called *simplicial*. The cluster which contains a simplicial is called *leaf cluster*.

An important relation between graphs and hypergraphs is given in [26]: A hypergraph is acyclic if and only if it can be considered to be the set of cliques of a triangulated graph (a graph is triangulated if every cycle of length greater than 4 has a chord).

In the Figure 1 one can see a) a triangulated graph, b) the corresponding acyclic hypergraph and c) the corresponding junction tree.

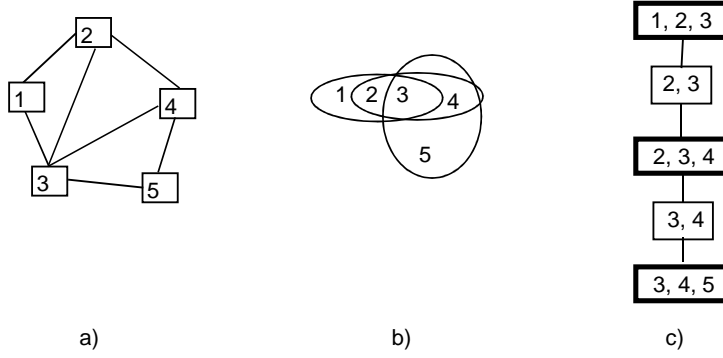


Fig. 1 a) Triangulated graph, b) The corresponding acyclic hypergraph, c) The corresponding junction tree which is a t -cherry junction tree

We consider the random vector $\mathbf{X} = (X_1, \dots, X_d)^T$, with the set of indices $V = \{1, \dots, d\}$. Roughly speaking a Markov network encodes the conditional independences between the random variables. The graph structure associated to a Markov network consists in the set of nodes V , and the set of edges $E = \{(i, j) \mid i, j \in V\}$.

We say that the probability distribution associated to a Markov network has the global Markov (GM) property [16] if in the graph $\forall A, B, C \subset V$ and C separates A and B in terms of graph then \mathbf{X}_A and \mathbf{X}_B are conditionally independent given \mathbf{X}_C , which means in terms of probabilities that

$$P(\mathbf{X}_{A \cup B \cup C}) = \frac{P(\mathbf{X}_{A \cup C}) P(\mathbf{X}_{B \cup C})}{P(\mathbf{X}_C)}.$$

The concept of *junction tree probability distribution* is related to the junction tree graph and to the global Markov property of the graph. A junction tree probability distribution is defined as a product and division of marginal probability distributions as follows:

$$P(\mathbf{X}) = \frac{\prod_{C \in \mathcal{C}} P(\mathbf{X}_C)}{\prod_{S \in \mathcal{S}} [P(\mathbf{X}_S)]^{\nu_S - 1}},$$

where \mathcal{C} is the set of clusters of the junction tree, \mathcal{S} is the set of separators, ν_S is the number of those clusters which contain the separator S . We emphasize here that the equalities written as $P(\mathbf{X}) = f(P(\mathbf{X}_K), K \in \mathcal{C})$, where $f : \Omega_{\mathbf{X}} \rightarrow R$ hold for any possible realization of \mathbf{X} .

Example 1 The probability distribution corresponding to Figure 1 is:

$$\begin{aligned} P(\mathbf{X}) &= \frac{P(\mathbf{X}_{\{1,2,3\}}) P(\mathbf{X}_{\{2,3,4\}}) P(\mathbf{X}_{\{3,4,5\}})}{P(\mathbf{X}_{\{2,3\}}) P(\mathbf{X}_{\{3,4\}})} \\ &= \frac{P(X_1, X_2, X_3) P(X_2, X_3, X_4) P(X_3, X_4, X_5)}{P(X_2, X_3) P(X_3, X_4)}. \end{aligned}$$

In our paper [33] we introduced a special kind of k -width junction tree, called k -th order t -cherry junction tree in order to approximate a joint probability distribution. The k -th order t -cherry junction tree probability distribution is assigned to the k -th order t -cherry tree, was introduced in [6], [7].

Definition 1 The recursive construction of the k -th order t -cherry tree:

- (i) The complete graph of $(k - 1)$ nodes from V represent the smallest k -th order t -cherry tree;
- (ii) By connecting a new vertex $i_k \in V$, with all $\{i_1, \dots, i_{k-1}\}$ vertices of a $(k - 1)$ - dimensional complete subgraph of the existing k -th order t -cherry tree, we obtain a new k -th order t -cherry tree. $\{\{i_k\} \{i_1, \dots, i_{k-1}\}\}$ is called k -th order hypercherry.
- (iii) A k -th order t -cherry tree can be obtained from (i) by successive application of (ii).

The k -th order t -cherry tree is a special triangulated (chordal or rigid circuit) graph therefore a junction tree structure is associated to it (see [26]).

Definition 2 ([33]) The k -th order t -cherry junction tree is defined in the following way:

- By using Definition 1 we construct a k -th order t -cherry tree over V .
- To each hypercherry $\{\{i_k\} \{i_1, \dots, i_{k-1}\}\}$ is assigned a cluster $\{i_1, \dots, i_{k-1}, i_k\}$ which represents a node of the junction tree and a separator $\{i_1, \dots, i_{k-1}\}$ which is an edge of the junction tree.

We denote by \mathcal{C}_{ch} , and \mathcal{S}_{ch} , the set of clusters and separators of the t -cherry junction tree.

Definition 3 ([33]) The probability distribution given by (1) and (2) are called t -cherry junction tree probability distribution

$$P_{\text{t-ch}}(\mathbf{X}) = \frac{\prod_{K \in \mathcal{C}_{\text{ch}}} P(\mathbf{X}_K)}{\prod_{S \in \mathcal{S}_{\text{ch}}} (P(\mathbf{X}_S))^{\nu_S - 1}} \quad (1)$$

in the discrete case and

$$P_{\text{t-ch}}(\mathbf{X}) = \frac{\prod_{K \in \mathcal{C}_{\text{ch}}} f_K(\mathbf{x}_K)}{\prod_{S \in \mathcal{S}_{\text{ch}}} (f_S(\mathbf{x}_S))^{\nu_S - 1}} \quad (2)$$

in the continuous case, where ν_S denotes the number of clusters which contain the separator S .

Remark 1 The marginal probability distributions and the density functions involved in the above formula are marginal probability distributions of $P(\mathbf{X})$.

Example 1 shows a 3-rd order t -cherry junction tree probability distribution.

In the following instead of probability distribution associated to a junction tree we will use shortly junction tree pd and similarly instead of k -th order t -cherry tree junction tree distribution we will use shortly k -th order t -cherry pd.

2.2 Copula, Regular-vine copula, junction tree copula and cherry-tree copula

Definition 4 A function $C : [0; 1]^d \rightarrow [0; 1]$ is called a d -dimensional copula if it satisfies the following conditions:

1. $C(u_1, \dots, u_d)$ is increasing in each component u_i ,
2. $C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_d) = 0$ for all $u_k \in [0; 1]$, $k \neq i$, $i = 1, \dots, n$,
3. $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ for all $u_i \in [0; 1]$, $i = 1, \dots, d$,
4. C is d -increasing, i.e for all $(u_{1,1}, \dots, u_{1,d})$ and $(u_{2,1}, \dots, u_{2,d})$ in $[0; 1]^d$ with $u_{1,i} < u_{2,i}$ for all i , we have

$$\sum_{i_1=1}^2 \cdots \sum_{i_d=1}^2 (-1)^{\sum_{j=1}^d i_j} C(u_{i_1,1}, \dots, u_{i_d,d}) \geq 0.$$

Due to Sklar's theorem if X_1, \dots, X_d are continuous random variables defined on a common probability space, with the univariate marginal cdf's $F_{X_i}(x_i)$ and the joint cdf $F_{X_1, \dots, X_d}(x_1, \dots, x_d)$ then there exists a unique copula function $C_{X_1, \dots, X_d}(u_1, \dots, u_d) : [0; 1]^d \rightarrow [0; 1]$ such that by the substitution $u_i = F_i(x_i)$, $i = 1, \dots, d$ we get

$$F_{X_1, \dots, X_d}(x_1, \dots, x_d) = C_{X_1, \dots, X_d}(F_1(x_1), \dots, F_d(x_d))$$

for all $(x_1, \dots, x_d)^T \in R^d$.

In the following we will use the vectorial notation $F_{\mathbf{X}_V}(\mathbf{x}_V) = C_{X_V}(\mathbf{u}_V)$, where $\mathbf{u}_V = (F_{X_{i_1}}(x_{i_1}), \dots, F_{X_{i_d}}(x_{i_d}))^T$.

It is known that

$$f_{X_{i_1}, \dots, X_{i_d}}(x_{i_1}, \dots, x_{i_d}) = c_{X_{i_1}, \dots, X_{i_d}}(F_{X_{i_1}}(x_{i_1}), \dots, F_{X_{i_d}}(x_{i_d})) \cdot \prod_{k=1}^d f_{X_{i_k}}(x_{i_k})$$

In vectorial notation this can be written as

$$f_{\mathbf{X}_V}(\mathbf{x}_V) = c_{\mathbf{X}_V}(\mathbf{u}_V) \cdot \prod_{i_k \in V} f_{X_{i_k}}(x_{i_k})$$

which can be written as

$$c_{\mathbf{X}_V}(\mathbf{u}_V) = \frac{f_{\mathbf{X}_V}(\mathbf{x}_V)}{\prod_{i_k \in V} f_{X_{i_k}}(x_{i_k})}.$$

The Regular-vine structures were introduced by T. Bedford and R. Cooke in [3], [4] and described in more detail in [25].

If it does not cause confusion, instead of $f_{\mathbf{X}_D}$ and $c_{\mathbf{X}_D}$ we will write f_D and c_D . We also introduce the following notations:

- $F_{i,j|D}$ – the conditional probability distribution function of X_i and X_j given \mathbf{X}_D ;
- $f_{i,j|D}$ – the conditional probability density function of X_i and X_j given \mathbf{X}_D ,
- $c_{i,j|D}$ – the conditional copula density corresponding to $f_{i,j|D}$,

where $D \subset V$; $i, j \in V \setminus D$.

According to the definition in [25]:

Definition 5 A *Regular-vine (R-vine)* on d variables consists first of a sequence of trees T_1, T_2, \dots, T_{d-1} with nodes N_i and edges E_i for $i = 1, \dots, d-1$, which satisfy the following conditions.

- T_1 has nodes $N_1 = \{1, \dots, d\}$ and edges E_1 .
- For $i = 2, \dots, d-1$ the tree T_i has nodes $N_i = E_{i-1}$.
- Two edges in tree T_i are joined in tree T_{i+1} only if they share a common node in tree T_i .

The last condition usually is referred to as *proximity condition*.

It is shown in [3] and [25] that the edges in an R-vine tree can be uniquely identified by two nodes, the conditioned nodes, and a set of conditioning nodes, i.e., edges are denoted by $e = j(e), k(e) | D(e)$ where $D(e)$ is the conditioning set. For a good overview see [12]. The next theorem which can be regarded as a central theorem of R-vines see [3] links the probability density function to the copulas assigned to the R-vine structure. In [3] it is shown that there exists a unique probability density assigned to the R-vine, in [4] it is shown that this probability distribution can be expressed as (3).

Theorem 1 The joint density of $\mathbf{X} = (X_1, \dots, X_d)$ is uniquely determined and given by:

$$f(x_1, \dots, x_d) = \left[\prod_{k=1}^d f_k(x_k) \right] \cdot \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e), k(e) | D(e)} \left(F_{j(e) | D(e)}(x_{j(e)} | \mathbf{x}_{D(e)}), F_{k(e) | D(e)}(x_{k(e)} | \mathbf{x}_{D(e)}) \right). \quad (3)$$

The arguments of the pair copulas are conditional distribution functions and can be evaluated using the following expression given by H. Joe [18]

$$\begin{aligned} & F_{j(e) | D(e)}(x_{j(e)} | \mathbf{x}_{D(e)}) \\ &= \frac{\partial C_{j(e), i | D(e) \setminus i}(F_{j(e) | D(e) \setminus i}(x_{j(e)} | \mathbf{x}_{D(e) \setminus i}), F_{i | D(e) \setminus i}(x_i | \mathbf{x}_{D(e) \setminus i}))}{\partial F_{i | D(e) \setminus i}(x_i | \mathbf{x}_{D(e) \setminus i})}, \end{aligned}$$

where $i \in D(e), j(e) \notin D(e)$.

We give now an other definition which is related to the k -th order t -cherry junction tree structure, see Definition 2, which is in fact a k -width order uniform hypertree.

Definition 6 The Regular-vine structure is given by a sequence of t -cherry junction trees T_1, T_2, \dots, T_{d-1} as follows

- T_1 is a regular tree on $V = \{1, \dots, d\}$, the set of edges $E_1 = \{e_i^1 = (l_i, m_i) | i = 1, \dots, d-1, l_i, m_i \in V\}$; The copula densities $c_{l_i, m_i}(F_{l_i}(x_{l_i}), F_{m_i}(x_{m_i}))$ are assigned to the edges of this tree.
- T_2 is the second order t -cherry junction tree on $V = \{1, \dots, d\}$, with the set of clusters $E_2 = \{e_i^2, i = 1, \dots, d-1 | e_i^2 = e_i^1\}$, $|e_i^1| = 2$; the copula densities

$$c_{a_{ij}^2, b_{ij}^2 | S_{ij}^2} \left(F_{a_{ij}^2 | S_{ij}^2}(x_{a_{ij}^2} | \mathbf{x}_{S_{ij}^2}), F_{b_{ij}^2 | S_{ij}^2}(x_{b_{ij}^2} | \mathbf{x}_{S_{ij}^2}) \right)$$

are assigned to each pair clusters e_i^2 and e_j^2 , which are linked in the junction tree T_2 , where:

$$\begin{aligned} S_{ij}^2 &= e_i^2 \cap e_j^2, \\ a_{ij}^2 &= e_i^2 - S_{ij}^2, \\ b_{ij}^2 &= e_j^2 - S_{ij}^2. \end{aligned}$$

- T_k is one of the possible k -th order t -cherry junction tree on $V = \{1, \dots, d\}$, with the set of clusters $E_k = \{e_i^k, i = 1, \dots, d - k + 1\}$, where each $e_i^k, |e_i^k| = k$ is obtained from the union of two linked clusters in the $(k - 1)$ -th order t -cherry junction tree T_{k-1} ; The copula densities

$$c_{a_{ij}^k, b_{ij}^k | S_{ij}^k} \left(F_{a_{ij}^k | S_{ij}^k} \left(x_{a_{ij}^k} | \mathbf{x}_{S_{ij}^k} \right), F_{b_{ij}^k | S_{ij}^k} \left(x_{b_{ij}^k} | \mathbf{x}_{S_{ij}^k} \right) \right)$$

are assigned to each pair of clusters e_i^k and e_j^k , which are linked in the T_k junction tree, where:

$$\begin{aligned} S^k &= e_i^k \cap e_j^k, \\ a_{ij}^k &= e_i^k - S_{ij}^k, \\ b_{ij}^k &= e_j^k - S_{ij}^k. \end{aligned}$$

Theorem 2 *The Regular-vine probability distribution associated to the R-vine structure given in Definition 6 can be expressed as:*

$$\begin{aligned} f(x_1, \dots, x_d) &= \left[\prod_{i=1}^d f_i(x_i) \right] \left[\prod_{i=1}^{d-1} c_{e_i^1} (F_{l_i} (x_{l_i}), F_{l_i} (x_{l_i})) \right] \\ &\cdot \prod_{i=2}^{d-1} \prod_{e \in E_i} c_{a_{ij}^k, b_{ij}^k | S_{ij}^k} \left(F_{a_{ij}^k | S_{ij}^k} \left(x_{a_{ij}^k} | \mathbf{x}_{S_{ij}^k} \right), F_{b_{ij}^k | S_{ij}^k} \left(x_{b_{ij}^k} | \mathbf{x}_{S_{ij}^k} \right) \right). \end{aligned}$$

For the following remark see [1], p. 186.

Remark 2 X_i and X_j are conditional independent given the set of variables $\mathbf{X}_A, A \subset V \setminus \{i, j\}$ if and only if

$$c_{ij|A} (F_{i|A} (x_i | \mathbf{x}_A), F_{j|A} (x_j | \mathbf{x}_A)) = 1.$$

The following theorem is an important consequence of Theorem 1.

Theorem 3 *If in an R-vine the conditional copula densities corresponding to the trees $T_k, T_{k+1}, \dots, T_{d-1}$ are all equal to 1 then there exists a joint probability distribution which can be expressed only with the conditional copula densities assigned to T_1, \dots, T_{k-1} :*

$$\begin{aligned} f(x_1, \dots, x_d) &= \left[\prod_{i=1}^d f_i(x_i) \right] \left[\prod_{i=1}^{d-1} c_{e_i^1} (F_{l_i} (x_{l_i}), F_{l_i} (x_{l_i})) \right] \\ &\cdot \prod_{i=2}^{k-1} \prod_{e \in E_i} c_{a_{ij}^k, b_{ij}^k | S_{ij}^k} \left(F_{a_{ij}^k | S_{ij}^k} \left(x_{a_{ij}^k} | \mathbf{x}_{S_{ij}^k} \right), F_{b_{ij}^k | S_{ij}^k} \left(x_{b_{ij}^k} | \mathbf{x}_{S_{ij}^k} \right) \right). \end{aligned}$$

The following definition of truncated vine at level k is given in [5].

Definition 7 A *pairwisely truncated R-vine at level k* (or truncated R-vine at level k) is a special R-vine copula with the property that all pair-copulas with conditioning set equal to, or larger than k , are set to bivariate independence copulas.

There arise the following questions. What special properties have the probability distribution, if we set to 1 the conditional copula densities associated to the trees T_k, \dots, T_{d-1} of its Regular vine? Which are the properties of the Markov network associated? We will answer these questions in Section 3 and Section 5.

The problem of finding the optimal truncation of the vine structure is formulated in [23] as follows: "If we assume that we can assign the independent copula to nodes of the vine with small absolute values of partial correlations, then this algorithm can be used to find an optimal truncation of a vine structure." Kurovicka defined as "best vine" the one whose nodes of the top trees (tree with most conditioning) correspond to the smallest absolute partial correlations. However small partial correlation result conditional independence only under restrictive assumption, so our approach deals with a more general case in Section 3.

In [21] we proved a theorem which connects the general junction tree probability distributions with the junction tree copulas. This theorem can be adapted to the t -cherry junction trees in the following way.

Theorem 4 *The copula density function associated to a junction tree probability distribution defined in Definition 3*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\prod_{K \in \mathcal{C}_{ch}} f_{\mathbf{X}_K}(\mathbf{x}_K)}{\prod_{S \in \mathcal{S}_{ch}} [f_{\mathbf{X}_S}(\mathbf{x}_S)]^{v_S-1}},$$

is given by

$$c_{\mathbf{X}}(\mathbf{u}_V) = \frac{\prod_{K \in \mathcal{C}_{ch}} c_{\mathbf{X}_K}(\mathbf{u}_K)}{\prod_{S \in \mathcal{S}_{ch}} [c_{\mathbf{X}_S}(\mathbf{u}_S)]^{v_S-1}}. \quad (4)$$

Definition 8 The copula density given by Formula (4) is called *t-cherry junction tree copula density* or simply *t-cherry copula*.

3 The characteristics of the Markov network associated to a continuous joint pd which can be expressed as a truncated R-vine

In this part we refer to Regular-vines as they are defined in Definition 6. First we illustrate the main ideas on an example.

The edge set of the first tree and the sequence of the t -cherry trees (in Figure 2) together with the copula densities determined by Definition 6 are the following:

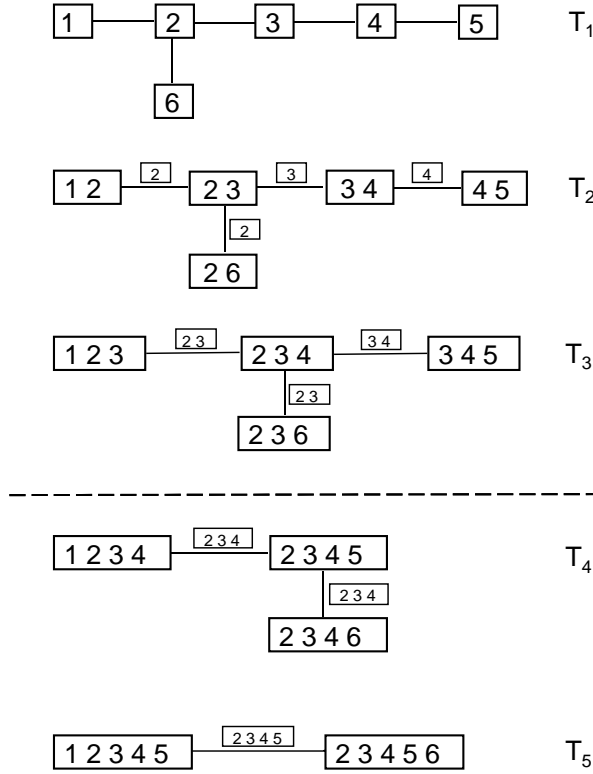


Fig. 2 Example for an R-vine structure on 6 variables using Definition 6

$$\begin{aligned}
T_1 : E_1 &= \{(1, 2), (2, 3), (2, 6), (3, 4), (4, 5)\}, \\
&\quad c_{1,2}, c_{2,3}, c_{2,6}, c_{3,4}, c_{4,5}; \\
T_2 : E_2 &= \{e_1^2 = (1, 2), e_2^2 = (2, 3), e_3^2 = (2, 6), e_4^2 = (3, 4), e_5^2 = (4, 5)\} \\
&\quad S_{1,2}^2 = e_1^2 \cap e_2^2 = \{2\}, \\
&\quad a_{1,2}^2 = e_1^2 - S_{1,2}^2 = \{1\}, b_{1,2}^2 = e_2^2 - S_{1,2}^2 = \{3\}, c_{a_{1,2}^2, b_{1,2}^2 | S_{1,2}^2} = c_{1,3|2} \\
&\quad S_{2,3}^2 = e_2^2 \cap e_3^2 = \{2\}, \\
&\quad a_{2,3}^2 = e_2^2 - S_{2,3}^2 = \{3\}, b_{2,3}^2 = e_3^2 - S_{2,3}^2 = \{6\}, c_{a_{2,3}^2, b_{2,3}^2 | S_{2,3}^2} = c_{3,6|2} \\
&\quad S_{2,4}^2 = e_2^2 \cap e_4^2 = \{3\}, \\
&\quad a_{2,4}^2 = e_2^2 - S_{2,4}^2 = \{2\}, b_{2,4}^2 = e_4^2 - S_{2,4}^2 = \{4\}, c_{a_{2,4}^2, b_{2,4}^2 | S_{2,4}^2} = c_{2,4|3} \\
&\quad S_{4,5}^2 = e_4^2 \cap e_5^2 = \{4\}, \\
&\quad a_{4,5}^2 = e_4^2 - S_{4,5}^2 = \{3\}, b_{4,5}^2 = e_5^2 - S_{4,5}^2 = \{5\}, c_{a_{4,5}^2, b_{4,5}^2 | S_{4,5}^2} = c_{3,5|4}; \\
T_3 : E_3 &= \{e_1^3 = (1, 2, 3), e_2^3 = (2, 3, 4), e_3^3 = (2, 3, 6), e_4^3 = (3, 4, 5)\} \\
&\quad S_{1,2}^3 = e_1^3 \cap e_2^3 = \{2, 3\}, \\
&\quad a_{1,2}^3 = e_1^3 - S_{1,2}^3 = \{1\}, b_{1,2}^3 = e_2^3 - S_{1,2}^3 = \{4\}, c_{a_{1,2}^3, b_{1,2}^3 | S_{1,2}^3} = c_{1,4|2,3} \\
&\quad S_{2,3}^3 = e_2^3 \cap e_3^3 = \{2, 3\}, \\
&\quad a_{2,3}^3 = e_2^3 - S_{2,3}^3 = \{4\}, b_{2,3}^3 = e_3^3 - S_{2,3}^3 = \{6\}, c_{a_{2,3}^3, b_{2,3}^3 | S_{2,3}^3} = c_{4,6|2,3} \\
&\quad S_{2,4}^3 = e_2^3 \cap e_4^3 = \{3, 4\}, \\
&\quad a_{2,4}^3 = e_2^3 - S_{2,4}^3 = \{2\}, b_{2,4}^3 = e_4^3 - S_{2,4}^3 = \{5\}, c_{a_{2,4}^3, b_{2,4}^3 | S_{2,4}^3} = c_{2,5|3,4}; \\
T_4 : E_4 &= \{e_1^4 = (1, 2, 3, 4), e_2^4 = (2, 3, 4, 5), e_3^4 = (2, 3, 4, 6)\} \\
&\quad S_{1,2}^4 = e_1^4 \cap e_2^4 = \{2, 3, 4\}, \\
&\quad a_{1,2}^4 = e_1^4 - S_{1,2}^4 = \{1\}, b_{1,2}^4 = e_2^4 - S_{1,2}^4 = \{5\}, c_{a_{1,2}^4, b_{1,2}^4 | S_{1,2}^4} = c_{1,5|2,3,4} \\
&\quad S_{2,3}^4 = e_2^4 \cap e_3^4 = \{2, 3, 4\}, \\
&\quad a_{2,3}^4 = e_2^4 - S_{2,3}^4 = \{5\}, b_{2,3}^4 = e_3^4 - S_{2,3}^4 = \{6\}, c_{a_{2,3}^4, b_{2,3}^4 | S_{2,3}^4} = c_{5,6|2,3,4} \\
T_5 : E_5 &= \{e_1^5 = (1, 2, 3, 4, 5), e_2^5 = (2, 3, 4, 5, 6)\} \\
&\quad S_{1,2}^5 = e_1^5 \cap e_2^5 = \{2, 3, 4, 5\}, \\
&\quad a_{1,2}^5 = e_1^5 - S_{1,2}^5 = \{1\}, b_{1,2}^5 = e_2^5 - S_{1,2}^5 = \{6\}, c_{a_{1,2}^5, b_{1,2}^5 | S_{1,2}^5} = c_{1,6|2,3,4,5}.
\end{aligned}$$

The joint probability density function of $\mathbf{X} = (X_1, \dots, X_6)$ can be expressed by Theorem 2 as follows:

$$\begin{aligned}
f(x_1, x_2, x_3, x_4, x_5, x_6) = & \left(\prod_{i=1}^6 f(x_i) \right) c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{2,3}(F_2(x_2), F_3(x_3)) \cdot c_{2,6}(F_2(x_2), F_6(x_6)) \\
& \cdot c_{3,4}(F_3(x_3), F_4(x_4)) \\
& \cdot c_{4,5}(F_4(x_4), F_5(x_5)) \\
& \cdot c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \\
& \cdot c_{3,6|2}(F_{3|2}(x_3|x_2), F_{6|2}(x_6|x_2)) \\
& \cdot c_{2,4|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)) \\
& \cdot c_{3,5|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)) \\
& \cdot c_{1,4|2,3}(F_{1|2,3}(x_1|x_2, x_3), F_{4|2,3}(x_4|x_2, x_3)) \\
& \cdot c_{4,6|2,3}(F_{4|2,3}(x_4|x_2, x_3), F_{6|2,3}(x_6|x_2, x_3)) \\
& \cdot c_{2,5|3,4}(F_{2|3,4}(x_2|x_3, x_4), F_{5|3,4}(x_5|x_3, x_4)) \\
& \cdot c_{1,5|2,3,4}(F_{1|2,3,4}(x_1|x_2, x_3, x_4), F_{5|2,3,4}(x_5|x_2, x_3, x_4)) \\
& \cdot c_{5,6|2,3,4}(F_{5|2,3,4}(x_5|x_2, x_3, x_4), F_{6|2,3,4}(x_6|x_2, x_3, x_4)) \\
& \cdot c_{1,6|2,3,4,5}(F_{1|2,3,4,5}(x_1|x_2, x_3, x_4, x_5), F_{6|2,3,4,5}(x_6|x_2, x_3, x_4, x_5))
\end{aligned}$$

In this part we regard the graph of the Markov network to be known. So let us suppose that the Markov network, which encodes the conditional probabilities between the random variables X_1, \dots, X_6 is given in Figure 3.

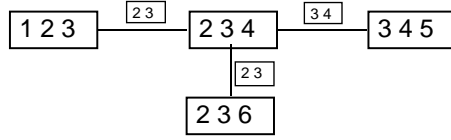


Fig. 3 3-rd order t -cherry junction tree

If the Markov network has the structure in Figure 3, then it is easy to identify the following conditional independences which are consequences of the Global Markov property:

$$\begin{aligned}
X_1 &\perp X_4 | X_2, X_3; & X_4 &\perp X_6 | X_2, X_3; & X_2 &\perp X_5 | X_3, X_4; \\
X_1 &\perp X_5 | X_2, X_3, X_4; & X_5 &\perp X_6 | X_2, X_3, X_4; \\
X_1 &\perp X_6 | X_2, X_3, X_4, X_5.
\end{aligned}$$

Based on the existence of these conditional independences the conditional copula densities associated to the trees T_3, T_4, T_5

$$\begin{aligned}
& c_{1,4|2,3}(F_{1|23}(x_1|x_2, x_3), F_{4|2,3}(x_4|x_2, x_3)), \\
& c_{4,6|2,3}(F_{1|23}(x_1|x_2, x_3), F_{4|2,3}(x_4|x_2, x_3)), \\
& c_{2,5|3,4}(F_{2,5|3,4}(x_2|x_3, x_4), F_{5|3,4}(x_5|x_3, x_4)), \\
& c_{1,5|2,3,4}(F_{1,5|2,3,4}(x_1|x_2, x_3, x_4), F_{5|2,3,4}(x_5|x_2, x_3, x_4)), \\
& c_{5,6|2,3,4}(F_{5|2,3,4}(x_5|x_2, x_3, x_4), F_{6|2,3,4}(x_6|x_2, x_3, x_4)), \\
& c_{1,6|2,3,4,5}(F_{1|2,3,4,5}(x_1|x_2, x_3, x_4, x_5), F_{6|2,3,4,5}(x_6|x_2, x_3, x_4, x_5)).
\end{aligned}$$

are all equal to 1. We can observe here that a Markov network of the form of a 3-rd order t -cherry tree (see Definition 1) can be expressed as an R-vine truncated at level 3.

This example suggests, that there are t -cherry tree probability distributions which can be represented as a truncated vines.

In the following we suppose the case when the set of separators of the k -th order t -cherry junction tree form a $(k - 1)$ -th order t -cherry junction tree. In this case we give an algorithm, which constructs a Regular-vine structure associated to a k -th order t -cherry tree probability distribution (see Definition 3).

Algorithm 1 Algorithm for obtaining from a t -cherry junction tree a truncated Regular-vine construction.

Input: A t -cherry tree structure, ie a set of clusters of size k , and the junction tree structure given by the separators.

Output: A Regular-vine truncated at level k .

We obtain recursively an $(m - 1)$ width t -cherry junction tree from a m - width t -cherry junction tree, for $m = k, \dots, 1$ as follows:

- 1. Step. The separators of the m -width t -cherry tree will be the clusters in the $(m - 1)$ -width t -cherry tree, which will be linked if between them is one cluster in the m -width t -cherry tree, and they are different.
- 2. Step. The leaf clusters, those clusters which contain a simplicial node, are transformed into $(m - 1)$ -width clusters, by deleting a node which is not simplicial. The cluster obtained in this way will be connected to one of the clusters obtained in Step 1, which was the separator linked to it in the m -width t -cherry tree junction tree.

Definition 9 The Regular-vine structure obtained from a t -cherry tree structure using Algorithm 1 is called *cherry-wine structure*.

Definition 10 The joint probability density assigned to a cherry-wine structure is called *cherry-wine density*, the corresponding copula density is called *cherry-wine copula density*.

Theorem 5 A t -cherry copula can be expressed as a cherry-wine copula in the following way:

$$\frac{\prod_{K \in \mathcal{C}_{ch}} c_K(\mathbf{u}_K)}{\prod_{S \in \mathcal{S}_{ch}} [c_S(\mathbf{u}_S)]^{v_S - 1}} = \left[\prod_{i=1}^{d-1} c_{e_i^1}(F_{l_i}(x_{l_i}), F_{l_i}(x_{l_i})) \right] \cdot \prod_{i=2}^{k-1} \prod_{e \in E_i} c_{a_{ij}^k, b_{ij}^k | S_{ij}^k} \left(F_{a_{ij}^k | S_{ij}^k} \left(x_{a_{ij}^k} | \mathbf{x}_{S_{ij}^k} \right), F_{b_{ij}^k | S_{ij}^k} \left(x_{b_{ij}^k} | \mathbf{x}_{S_{ij}^k} \right) \right).$$

Example in Figure 4 shows how to apply Algorithm 1 to a given 3-rd order t -cherry junction tree to obtain a cherry-wine structure.

The cherry wine probability distribution assigned to the 3-rd order cherry-wine structure in Figure 4 is:

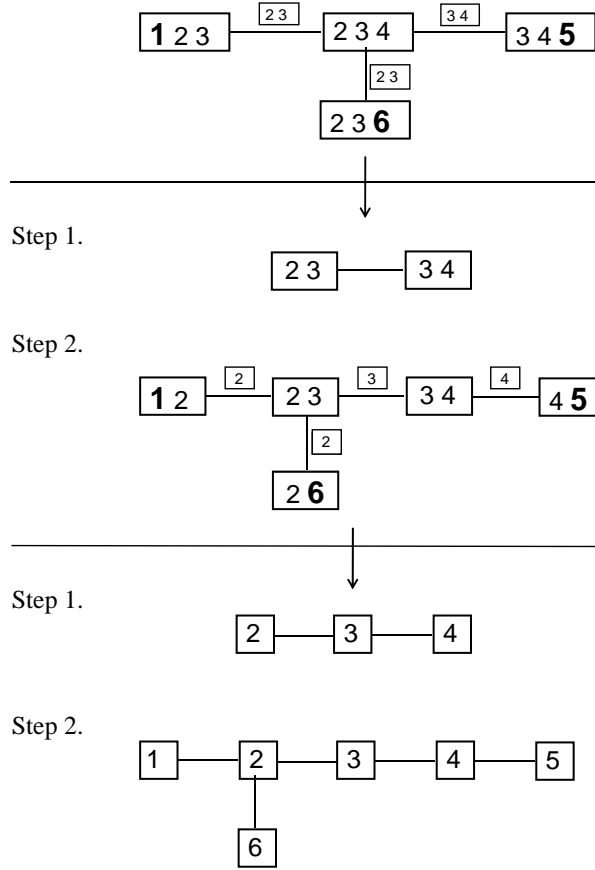


Fig. 4 Application of Algorithm 1 to a 3-rd order t -cherry junction tree in order to obtain a 3-rd order cherry-wine structure

$$\begin{aligned}
 f(x_1, x_2, x_3, x_4, x_5, x_6) &= \\
 &= \left(\prod_{i=1}^6 f(x_i) \right) \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{2,3}(F_2(x_2), F_3(x_3)) \\
 &\cdot c_{2,6}(F_2(x_2), F_6(x_6)) \cdot c_{3,4}(F_3(x_3), F_4(x_4)) \cdot c_{4,5}(F_4(x_4), F_5(x_5)) \\
 &\cdot c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot c_{3,6|2}(F_{3|2}(x_3|x_2), F_{6|2}(x_6|x_2)) \\
 &\cdot c_{2,4|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)) \cdot c_{3,5|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4))
 \end{aligned}$$

Remark 3 Applying Algorithm 1 can result more cherry-wine structures since in Step 2 we can proceed in different directions.

Starting from the 3-rd order t -cherry junction tree given in Figure 3 we can obtain $2^{\#\text{leaf clusters}} = 8$ 2-nd order t -cherry trees. In the last step there is only one possibility to construct the first tree. We emphasize here that, if the Markov network has the 3-rd order t -cherry tree structure in Figure 3, than from the 23,040 possible R-vines (see [13], [30]) remain only 8.

The question, which arises here is whether the 3-rd order t -cherry junction tree is not a very special structure.

We proved in [33] the following theorem by a constructive method:

Theorem 6 *Any k -width junction tree probability distribution can be expressed as a k -th order t -cherry tree probability distribution.*

Remark 4 There exists more expressions for the t -cherry probability distribution, but much smaller number than Regular-vines, and has the advantage of exploiting the conditional independences.

4 A model selection for special R-vines called cherry-wines

Now we suppose to have a sample data set. Starting from this dataset we want to find a good fitting probability distribution. The main idea is fitting the copula function and the marginal probability distribution separately. Using pair-vine constructions we will express the joint density function only by marginal distributions and bivariate (pair)-copulas. First we will search for a good fitting regular vine structure. As it is shown in [30] the number of possible regular vines grows exponentially with the number of variables. So the basic idea is searching through truncated R-vine copulas at a given level k .

Full inference for pair-copula decomposition should in principle consider three elements [1]:

- The selection of a specific factorization;
- The choice of pair-copula types;
- The estimation of parameters of the chosen pair-copulas.

This paper is concerned with finding of factorization which exploits some of the conditional independences between the random variables.

There are many papers dealing with selecting specific Regular-vines as C-vine or D-vine see for example [1].

The main idea of our approach is finding a t -cherry copula and then transforming it by Algorithm 1 into a cherry-wine copula, which depends just on pair-copulas. So we will start at a given level k , search for the best fitting t -cherry copula to the sample data and find then the factorization which results the chosen k -th order t -cherry tree.

4.1 The Sample derived copula

The empirical probability distribution of the sample data is a discrete multivariate probability distribution. If this data is drawn i.i.d from a continuous joint probability distribution all realizations are different vectors. So the joint probability distribution is uniform. The range of each random variable is equal to the sample size N .

As it is shown in [21] we first make a partition of the range of each random variable involved. The intervals obtained contain the same number of data. We introduced a special type of copula called *sample derived copula*.

We denote the set of the values of X_i in the sample by Λ_i . This set contains N values, for each random variable. The theoretical range of the continuous random variable X_i will be denoted by $\bar{\Lambda}_i$. For every i we denote by $\lambda_i^m = \min \bar{\Lambda}_i \in R$ and by $\lambda_i^M = \max \bar{\Lambda}_i \in R$. We suppose for simplicity that $\min \bar{\Lambda}_i \neq \min \Lambda_i$ and $\max \bar{\Lambda}_i \neq \max \Lambda_i$. For each random variable X_i we define a partition of Λ_i by $\mathcal{P}_i = \left\{ x_0^{p_i} = \lambda_i^m, x_1^{p_i}, \dots, x_{m_i-1}^{p_i}, x_{m_i}^{p_i} = \lambda_i^M \right\}$ with the following properties:

- For each random variable X_i , each interval $(x_{j-1}^{p_i}; x_j^{p_i}]$, $j = 1, \dots, m_i$ contains a given $n_i = \frac{N}{m_i} \in N$ number of values from the set Λ_i .
- Each $x_j^{p_i} \in \Lambda_i$, $j = 1, \dots, m_i - 1$.

The partition with the above properties will be called uniform partition. We denote by \mathcal{P} the set of partitions $\{\mathcal{P}_1, \dots, \mathcal{P}_n\}$.

Let be \widetilde{X}_i the categorical random variable associated to the random variable X_i :

$$P(\widetilde{X}_i \in (x_{j-1}^{p_i}; x_j^{p_i}]) = \frac{1}{m_i}, j = 1, \dots, m_i.$$

We assign to each $x^i \in (x_{j-1}^{p_i}; x_j^{p_i}]$ the number $u_j^i = \frac{j}{m_i}$, $j = 0, \dots, m_i$. Obviously $u_0^i = 0$ and $u_{m_i}^i = 1$. Let $\tilde{\Lambda}_i = \{u_j^i | j = 0, \dots, m_i\}$. So we can define the following discrete uniform random variables:

$$\tilde{U}_i = \begin{pmatrix} u_0^i & u_1^i & \dots & u_{m_i-1}^i & u_{m_i}^i \\ 0 & \frac{1}{m_i} & \dots & \frac{1}{m_i} & \frac{1}{m_i} \end{pmatrix}, i = 1, \dots, d.$$

Now we transform the sample using the above assignment. We denote the transformed sample by \mathcal{T} .

Definition 11 The function $\tilde{c} : \prod_{i=1}^d \tilde{\Lambda}_i \rightarrow R$ defined by

$$(u_{k_1}^1, \dots, u_{k_d}^d) \mapsto \tilde{c}(u_{k_1}^1, \dots, u_{k_d}^d) = \frac{\#\{(u_{k_1}^1, \dots, u_{k_d}^d) \in \mathcal{T}\}}{N}, k_i = 0, \dots, m_i$$

will be called *sample derivated copula density*.

In Remark 6 of the paper [21] we proved also, that partitioning in this way the information content of the joint probability distribution depends just on the sample derivated copula.

The sample derivated copula can be treated as a discrete multivariate probability distribution. One of its advantages is that the range of the variables involved are significantly decreased.

Now using the greedy Szántai-Kovács algorithm introduced in [35] we find the k -th order t -cherry copula. The goodness of fit to the data is quantified by Kullback-Leibler divergence. We emphasize here that finding the best fitting t -cherry copula is an NP-hard problem for $k > 2$, but there are cases, when the greedy algorithm finds the optimal solution, see [35].

4.2 The Szántai-Kovács greedy algorithm

We present here the algorithm introduced in [35].

The following theorem regarded to discrete probability distributions given in [21].

In [33] the authors give the following theorem.

Theorem 7 *The Kullback-Leibler divergence between the true $P(\mathbf{X})$ and the approximation given by the k -width junction tree probability distribution $P(\mathbf{X}_J)$, determined by the set of clusters \mathcal{C} and the set of separators \mathcal{S} is :*

$$KL(P(\mathbf{X}), P_J(\mathbf{X})) = -H(\mathbf{X}) - \left(\sum_{C \in \mathcal{C}} I(\mathbf{X}_C) - \sum_{S \in \mathcal{S}} (\nu_S - 1) I(\mathbf{X}_S) \right) + \sum_{i=1}^d H(X_i), \quad (5)$$

where $I(\mathbf{X}_C) = \sum_{i \in C} H(X_i) - H(\mathbf{X}_C)$ represents the information content of the random vector \mathbf{X}_C and similarly $I(\mathbf{X}_S) = \sum_{i \in S} H(X_i) - H(\mathbf{X}_S)$ represents the information content of the random vector \mathbf{X}_S .

In Formula (5) $-H(\mathbf{X}) + \sum_{i=1}^d H(X_i) = I(\mathbf{X})$ is independent from the structure of the junction tree. It is easy to see that minimizing the Kullback-Leibler divergence means maximizing $\sum_{C \in \mathcal{C}} I(\mathbf{X}_C) - \sum_{S \in \mathcal{S}} (\nu_S - 1) I(\mathbf{X}_S)$. We call this sum as *weight of the junction tree pd*. As larger this weight is, as better fits the approximation associated to the junction tree pd to the true probability distribution. It is well known that $KL = 0$ if $P(\mathbf{X}) = P_J(\mathbf{X})$.

Definition 12 We define the following concepts:

- the search space:

$$E = \{ \chi_{i_k(i_1, \dots, i_{k-1})} = \{ \{X_{i_k}\}, \{X_{i_1}, \dots, X_{i_{k-1}}\} \} \mid X_{i_1}, \dots, X_{i_{k-1}}, X_{i_k} \in X \},$$

- the independence set:

$$\mathcal{F} = \phi \cup \{t\text{-cherry junction tree structure}\},$$

- the weight function:

$$w : E \rightarrow R \quad w(\chi_{i_k(i_1, \dots, i_{k-1})}) = I(X_{i_1}, \dots, X_{i_{k-1}}, X_{i_k}) - I(X_{i_1}, \dots, X_{i_{k-1}}).$$

Algorithm 2 Szántai-Kovács's greedy algorithm.

Input: Elements of E and their weights which can be calculated based on the k -th order marginal probability distributions.

Output: set A which contains the clusters of the k -th order t -cherry junction tree pd and the weight of the k -th order t -cherry junction tree pd.

The algorithm:

$A := \phi$

Sort E into monotonically decreasing order by weight w ;

Choose $x = \arg \max_{x \in E} (w(x))$;

```

    let  $A := A \cup \{x\}$ ;    $E := E \setminus \{x\}$ ;    $w := I(x)$ ;
Do for each  $x \in E$  taken in monotonically decreasing order
    if  $A \cup \{x\} \in \mathcal{F}$  then let  $A := A \cup \{x\}$ ;    $E := E \setminus \{x\}$ ;    $w := w + w(x)$ ;
    if the union of subsets of  $A$  is  $X$ , then Stop;
    else take the next element of  $E$ .

```

4.3 Building the cherry-wine associated to the t -cherry tree.

We calculate the k -th order marginal pd from the sample derivated copula. Using their information content we can define the weights of the elements of the search space E .

Applying Szántai-Kovács's algorithm we obtain a good fitting t -cherry tree copula.

We assign to this the k -th order t -cherry tree the T_k tree of a regular vine. Applying now Algorithm 1 we can find the corresponding cherry-wine structure, and using this the expression of the cherry-wine copula density expressed by pair-copulas.

Now comes the next step the choice of pair-copula types and the estimation of parameters. For choosing pair copulas we have a large amount of copula-families, with different properties, tail-dependencies see in [18], [14] and [31].

5 Properties of the best fitting cherry-wine probability density, and cherry-wine copula density

In this section we discuss the properties of the best fitting cherry-wine probability density and corresponding copula density, which are associated to an R-vine truncated at level k from a theoretical point of view.

We will use the following notations:

- $f_V(\mathbf{x}_V)$ denotes the joint probability density of \mathbf{X}_V , $f_K(\mathbf{x}_K)$ is the marginal density of $f_V(\mathbf{x}_V)$, where $K \subset V$.
- $c_V(\mathbf{u}_V)$ denotes the joint copula density associated to the joint probability density $f_V(\mathbf{x}_V)$, $c_K(\mathbf{u}_K)$ is its marginal density which is the copula density corresponding to $f_K(\mathbf{x}_K)$
- $\hat{f}_{V_{CS}}$ denotes the joint k -th order cherry-wine density, associated to a k -th order t -cherry junction tree with cluster \mathcal{C} and separator set \mathcal{S} , given by:

$$\hat{f}_{V_{CS}}(\mathbf{x}_V) = \frac{\prod_{K \in \mathcal{C}_{ch}} f_K(\mathbf{x}_K)}{\prod_{S \in \mathcal{S}_{ch}} (f_S(\mathbf{x}_S))^{\nu_S - 1}}, \quad (6)$$

where ν_S is the number of clusters which contain S .

Theorem 8 *The Kullback-Leibler divergence between $f_V(\mathbf{x}_V)$ and the approximating probability density assigned to the cherry-wine $\hat{f}_{V_{CS}}$, is given by the formula:*

$$KL\left(\widehat{f}_{V_{CS}}(\mathbf{x}_V), f_V(\mathbf{x}_V)\right) = I(\mathbf{X}_V) - \left[\sum_{K \in \mathcal{C}_{ch}} I(\mathbf{X}_K) - \sum_{S \in \mathcal{S}_{ch}} (\nu_S - 1) I(\mathbf{X}_S) \right]. \quad (7)$$

Proof

$$\begin{aligned} KL\left(\widehat{f}_{V_{CS}}(\mathbf{x}_V), f_V(\mathbf{x}_V)\right) &= \int_{R^d} f_V(\mathbf{x}) \log_2 \frac{f_V(\mathbf{x})}{\widehat{f}_{V_{CS}}(\mathbf{x})} d\mathbf{x} \\ &= \int_{R^d} f_V(\mathbf{x}) \log_2 f_V(\mathbf{x}) d\mathbf{x} - \int_{R^d} f_V(\mathbf{x}) \log_2 \widehat{f}_{V_{CS}}(\mathbf{x}) d\mathbf{x} \\ &= -H(\mathbf{X}) - \int_{R^d} f_V(\mathbf{x}) \log_2 \frac{\prod_{K \in \mathcal{C}} f_K(\mathbf{x}_K)}{\prod_{S \in \mathcal{S}} (f_S(\mathbf{x}_S))^{\nu_S - 1}} d\mathbf{x} \\ &= -H(\mathbf{X}) - \int_{R^d} f_V(\mathbf{x}) \left[\log_2 \prod_{K \in \mathcal{C}} f_K(\mathbf{x}_K) - \log_2 \prod_{S \in \mathcal{S}} (f_S(\mathbf{x}_S))^{\nu_S - 1} \right] d\mathbf{x} \\ &= -H(\mathbf{X}) - \int_{R^d} f_V(\mathbf{x}) \log_2 \prod_{K \in \mathcal{C}} f_K(\mathbf{x}_K) d\mathbf{x} + \int_{R^d} f_V(\mathbf{x}) \log_2 \prod_{S \in \mathcal{S}} (f_S(\mathbf{x}_S))^{\nu_S - 1} d\mathbf{x}. \end{aligned}$$

Since $\bigcup_{K \in \mathcal{C}} K = V$ and each variable belongs once more to the clusters than to the separators, by adding and subtracting

$$\int_{R^d} f_V(\mathbf{x}) \log_2 \prod_{K \in \mathcal{C}} \prod_{i \in K} f_i(x_i) d\mathbf{x}$$

we obtain

$$\begin{aligned} KL\left(\widehat{f}_{V_{CS}}(\mathbf{x}), f_V(\mathbf{x})\right) &= -H(\mathbf{X}) - \int_{R^d} f_V(\mathbf{x}) \log_2 \frac{\prod_{K \in \mathcal{C}} f_K(\mathbf{x}_K)}{\prod_{K \in \mathcal{C}} \prod_{i \in K} f_i(x_i)} d\mathbf{x} \\ &+ \int_{R^d} f_V(\mathbf{x}) \log_2 \frac{\prod_{S \in \mathcal{S}} [f_S(\mathbf{x}_S)]^{\nu_S - 1}}{\prod_{S \in \mathcal{S}} \left[\prod_{i \in K} f_i(x_i) \right]^{\nu_S - 1}} d\mathbf{x} - \int_{R^d} f_V(\mathbf{x}) \log_2 \prod_{i=1}^d f_i(x_i) d\mathbf{x} \\ &= -H(\mathbf{X}) - \int_{R^d} f_V(\mathbf{x}) \sum_{K \in \mathcal{C}} \log_2 \frac{f_K(\mathbf{x}_K)}{\prod_{i \in K} f_i(x_i)} d\mathbf{x} \\ &+ \int_{R^d} f_V(\mathbf{x}) \sum_{S \in \mathcal{S}} \log_2 \frac{[f_S(\mathbf{x}_S)]^{\nu_S - 1}}{\left[\prod_{i \in S} f_i(x_i) \right]^{\nu_S - 1}} d\mathbf{x} - \int_{R^d} f_V(\mathbf{x}) \sum_{i=1}^d \log_2 f_i(x_i) d\mathbf{x} \end{aligned}$$

Since $f_K(\mathbf{x}_K)$, $f_S(\mathbf{x}_S)$, $f_i(x_i)$ are consistent marginals of $f_V(\mathbf{x})$ we have the following relations:

$$\begin{aligned} &\int_{R^d} f_V(\mathbf{x}) \sum_{K \in \mathcal{C}} \log_2 \frac{f_K(\mathbf{x}_K)}{\prod_{i \in K} f_i(x_i)} d\mathbf{x} \\ &= \sum_{K \in \mathcal{C}} \int_{R^k} f_K(\mathbf{x}) \log_2 \frac{f_K(\mathbf{x}_K)}{\prod_{i \in K} f_i(x_i)} d\mathbf{x}_K = \sum_{K \in \mathcal{C}} I(\mathbf{X}_K) \end{aligned} \quad (8)$$

$$\begin{aligned} & \int_{R^d} f_V(\mathbf{x}) \sum_{S \in \mathcal{S}} \log_2 \frac{[f_S(\mathbf{x}_S)]^{\nu_S-1}}{\prod_{i \in S} f_i(x_i)} d\mathbf{x} \\ &= \sum_{S \in \mathcal{S}} (\nu_S - 1) \int_{R^{k-1}} f_S(\mathbf{x}) \log_2 \frac{[f_S(\mathbf{x}_S)]}{\prod_{i \in S} f_i(x_i)} d\mathbf{x}_S = \sum_{S \in \mathcal{C}} (\nu_S - 1) I(\mathbf{X}_S) \end{aligned} \quad (9)$$

$$\begin{aligned} & - \int_{R^d} f_V(\mathbf{x}) \sum_{i=1}^d \log_2 f_i(x_i) \\ &= \sum_{i=1}^d - \int_{-\infty}^{\infty} f_i(\mathbf{x}_i) \log_2 f_i(x_i) dx_i = \sum_{i=1}^d H(X_i) \end{aligned} \quad (10)$$

where $I(\mathbf{X}_K)$, $I(\mathbf{X}_S)$ are the information contents (see [10]) of the \mathbf{X}_K and \mathbf{X}_S corresponding to the index set $K \in \mathcal{C}$ and $S \in \mathcal{S}$.

Taking into account relations (8), (9) and (10) we obtain:

$$KL(\hat{f}_{V_{CS}}(\mathbf{x}), f_V(\mathbf{x})) = \sum_{i=1}^d H(X_i) - H(\mathbf{X}) - \left[\sum_{K \in \mathcal{C}_{ch}} I(\mathbf{X}_K) - \sum_{S \in \mathcal{S}_{ch}} (\nu_S - 1) I(\mathbf{X}_S) \right]$$

As we know that

$$\sum_{i=1}^d H(X_i) - H(\mathbf{X}) = I(\mathbf{X})$$

we obtained formula (7) and this proves the theorem.

It is easy to see that the difference $I(\mathbf{X})$ do not depend on the structure of the junction tree. A consequence of Theorem 8 is the following remark.

Remark 5 The probability density $\hat{f}_{V_{CS}}$ of the form (6), which is the best fitting cherry-wine to the real probability density f_V over all possible truncated R-vines at level k maximizes the following difference

$$\sum_{K \in \mathcal{C}_{ch}} I(\mathbf{X}_K) - \sum_{S \in \mathcal{S}_{ch}} (\nu_S - 1) I(\mathbf{X}_S).$$

Now we make some observation on the corresponding copula densities.

For two variables it was shown (see [8] and [27]) that:

$$I(X, Y) = \int_{[0;1]^2} c(u, v) \log_2 c(u, v) du dv$$

which means that information content is equivalent with "copula entropy" concept introduced in [27].

Generalizing this for the variables involved in the sets K and S we have:

$$\begin{aligned} I(\mathbf{X}_K) &= \int_{[0;1]^k} c(\mathbf{u}_{\mathbf{X}_K}) \log_2 c(\mathbf{u}_{\mathbf{X}_K}) d\mathbf{u}_{\mathbf{X}_K} = -H(c_{\mathbf{X}_K}) \\ I(\mathbf{X}_S) &= \int_{[0;1]^{k-1}} c(\mathbf{u}_{\mathbf{X}_S}) \log_2 c(\mathbf{u}_{\mathbf{X}_S}) d\mathbf{u}_{\mathbf{X}_S} = -H(c_{\mathbf{X}_S}) \end{aligned}$$

Using the above assertions in Theorem 8 the Kullback-Leibler divergence can be expressed by means of copula entropies:

$$KL\left(\hat{f}_{V_{CS}}(\mathbf{x}_V), f_V(\mathbf{x}_V)\right) = H(c_{\mathbf{X}_V}) + \left[\sum_{K \in \mathcal{C}_{ch}} H(c_{\mathbf{X}_K}) - \sum_{S \in \mathcal{S}_{ch}} (\nu_S - 1) H(c_{\mathbf{X}_S}) \right].$$

Remark 6 The cherry-wine copula density $\hat{c}_{V_{CS}}$ associated to the best fitting cherry-wine probability density $\hat{f}_{V_{CS}}$ minimizes the following difference over all possible truncated R-vines at level k :

$$\sum_{K \in \mathcal{C}_{ch}} H(c_{\mathbf{X}_K}) - \sum_{S \in \mathcal{S}_{ch}} (\nu_S - 1) H(c_{\mathbf{X}_S}).$$

6 Conclusion

In this paper we gave an alternative definition of Regular-vines using the concept of t -cherry junction tree. We introduced the cherry-wine structure (a truncated R-vine assigned to a t -cherry probability distribution). We gave an algorithm for constructing a truncated R-vine at level k starting from special k -th order t -cherry junction trees. The problem of inference was also discussed. We developed a method for obtaining a good factorization (which exploits conditional independences) starting from a sample data. In the last section we discussed some theoretical properties of the best fitting truncated R-vine. In future we are planning to extend our algorithm to the general case.

References

1. K. Aas, C. Czado, A. Frigessi, and H. Bakken, Pair-copula constructions of multiple dependence, *Insur. Math. Econ.*, 44, 182–198, (2009)
2. A. Bauer, C. Czado and T. Klein, Pair-copula construction for non-Gaussian DAG models, submitted, (2011)
3. T. Bedford and R. Cooke, Probability density decomposition for conditionally dependent random variables modeled by vines, *Ann. Math. Artif. Intell.*, 32, 245–268, (2001)
4. T. Bedford and R. Cooke, Vines – a new graphical model for dependent random variables, *Ann. Stat.*, 30(4), 1021–1068, (2002)
5. E.C. Brechmann, C. Czado and K. Aas, Truncated regular vines in high dimensions with applications to financial data, Submitted for publication, [www-m4.ma.tum.de/Papers/Brechmann/vinetrunc.pdf](http://www.m4.ma.tum.de/Papers/Brechmann/vinetrunc.pdf) (2010)
6. J. Bukszár and A. Prékopa, Probability Bounds with Cherry Trees, *Mathematics of Operational Research* 26, 174–192, (2001)
7. J. Bukszár and T. Szántai, Probability Bounds given by hypercherry trees, *Optimization Methods and Software*, 17, 409–422, (2002)
8. R.S. Calsaverini and R. Vicente, An information theoretic approach to statistical dependence: Copula information, *arXiv:0911.4207v1*, (2009)
9. C.K. Chow and C.N. Liu, Approximating Discrete Probability Distribution with Dependence Tree, *IEEE Transactions on Information Theory*, 14, 462–467, (1968)
10. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Interscience, New York, (1991)
11. R.G. Cowell, P.A. Dawid, S.L. Lauritzen and D.J. Spiegelhalter, *Probabilistic Networks and Expert Systems (Information Science and Statistics)*, Springer, Heidelberg, (2003)

12. C. Czado, Pair-copula constructions of multivariate copulas, In: P. Jaworski, F. Durante, W. Hrdle and T. Rychlik (Eds.), *Copula Theory and Its Applications*, Berlin, Springer, (2010)
13. J.F. Dissmann, Statistical Inference for Regular Vines and Application, Thesis, Technical University of Munich, Center of Mathematics, (2010)
14. C. Genest and L.P. Rivest, Statistical inference procedures for bivariate Archimedean copulas, *J. Am. Stat. Assoc.*, 88(423), 1034–1043, (1993)
15. I.H. Haff, K. Aas and A. Frigessi, On the simplified pair-copula construction – simply useful or too simplistic? Technical Report, Norwegian Computing Center, Oslo (2009)
16. J.M. Hammersley, Markov fields on finite graphs and lattices, not published, (1971)
17. A. Hanea, D. Kurowicka and R. Cooke, Hybrid method for quantifying and analyzing Bayesian belief networks, *Qual. Reliab. Eng.*, 22, 708–729, (2006)
18. H. Joe, *Multivariate Models and Dependence Concepts*, Chapman & Hall, London, (1997)
19. D. Karger and N. Srebro, Learning networks: Maximum likelihood bounded tree-width graphs, *SODA-01*
20. E. Kovács and T. Szántai, On the approximation of discrete multivariate probability distribution using the new concept of t -cherry junction tree, *Lecture Notes in Economics and Mathematical Systems*, 633, Proceedings of the IFIP/IIASA/GAMM Workshop on Coping with Uncertainty, Robust Solutions, 2008, IIASA, Laxenburg, 39–56, (2010)
21. E. Kovács and T. Szántai, Multivariate copula expressed by lower dimensional copulas, <http://arxiv.org/abs/1009.2898>, (2010)
22. S. Kullback, *Information Theory and Statistics*, Wiley and Sons, New York, (1959)
23. D. Kurowicka, Optimal truncation of vines, in: D. Kurowicka and H. Joe (eds) *Dependence-Modeling – Handbook on Vine Copulas*, World Scientific Publishing, Singapore, (2010)
24. D. Kurowicka and R. Cooke, The vine copula method for representing high dimensional dependent distributions: Application to continuous belief nets, *Proceedings of the 2002 Winter Simulation Conference*, 270–278, (2002)
25. D. Kurowicka and R. M. Cooke, *Uncertainty Analysis with High Dimensional Dependence Modelling*, Chichester, John Wiley, (2006)
26. S.L. Lauritzen and D.J. Spiegelhalter, Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems, *J.R. Statist. Soc. B*, 50, 157–227, (1988)
27. J. Ma and Z. Sun, Mutual information is copula entropy, *arXiv: 0808.0845v1*, (2008)
28. A. Min and C. Czado, Bayesian inference for multivariate copulas using pair-copula constructions, *J. Financ. Econom.*, to appear, preprint available under: <http://www-m4.ma.tum.de/Papers/index.html>, (2010)
29. A. Min and C. Czado, Bayesian model selection for multivariate copulas using pair-copula constructions, Preprint, (2009)
30. O. Morales-Napoles, R. Cooke and D. Kurowicka, About the number of vines and regular vines on n nodes, *Discrete Appl. Math.*, Submitted (2010)
31. R. Nelsen, *An Introduction to Copulas*, Springer, New York, (1999)
32. J. Pearl, *Probabilistic reasoning in intelligent systems*, CA: Morgan Kaufman, Palo Alto, (1988)
33. Szántai, T. and E. Kovács, Hypergraphs as a mean of discovering the dependence structure of a discrete multivariate probability distribution, *Proc. Conference APplied mathematical programming and MODelling (APMOD), 2008*, Bratislava, 27–31 May 2008, *Annals of Operations Research*, to appear.
34. T. Szántai and E. Kovács, Application of t -cherry junction trees in pattern recognition, *Broad Research in Artificial Intelligence and Neuroscience (BRAIN)*, Special Issue on Complexity in Sciences and Artificial Intelligence, Eds. B. Iantovics, D. Radoiu, M. Marusteri and M. Dehmer, 40–45, (2010)
35. T. Szántai and E. Kovács, Discovering a junction tree behind a Markov network by a greedy algorithm, <http://arxiv.org/abs/1104.2762>, (2010)
36. R.E. Tarjan and M. Yannakakis, Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs and selectively reduce acyclic hypergraphs, *SIAM J. Comp.*, 13, 566–579, (1984)
37. M. Yannakakis, Computing the minimum Fill in is NP-complete, *SIAM Journal Alg. Disc. Math.*, 2, 77–79, (1981)